ENERGY EFFICIENT VIRTUAL MACHINE CONSOLIDATION FOR CLOUD ENVIRONMENT

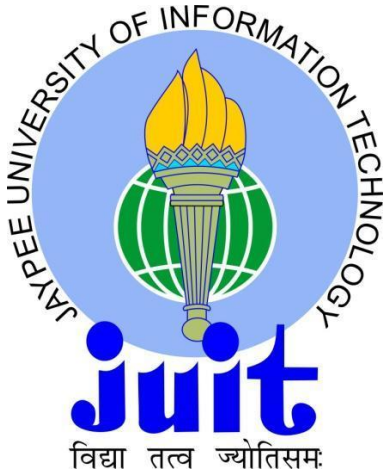
Synopsis submitted in fulfillment of the requirement for the Degree of

DOCTOR OF PHILOSOPHY

BY

OSHIN SHARMA

Enrollment No. 146202



Under the supervision of

Dr. Hemraj Saini

Department of Computer Science and Engineering

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

Waknaghat, Solan-173234, Himachal Pradesh, INDIA

September, 2017

ABSTRACT

Cloud computing offers an on-demand computing model which has revolutionized the working environment of IT industry and thus, the trend of cloud computing is increasing with the passage of time. Out of many different services of cloud environment, IaaS (Infrastructure as a service) is one of the dominant service which has given the birth to pay per use model and on the basis of this model, service providers provide services to the users. To satisfy the requirements of cloud users, this environment requires large number of physical devices and, it needs a huge amount of electricity to power and cool down the various electric components present inside these devices. It has been observed that the electricity consumption of data centers is 1.1% to 1.5% of overall electricity consumed in the world and it is growing by 12% every year. This high consumption of electricity results in the emissions of high carbon dioxide and therefore, it is very important to minimize the consumption of energy for the sake of environmental sustainability. Energy consumption of data center can be minimized by using lesser number of resources and improving the utilization of these resources. Moreover, the energy consumption can be minimized by deactivating and reactivating the number of physical hosts present inside data center.

This thesis presents new model, algorithms and heuristics for all different steps of dynamic virtual machine consolidation in Infrastructure as a service architecture that meets the requirements of SLA agreement and deals with energy-performance trade-off. The process of virtual machine (VM) consolidation has been chosen for minimizing the energy consumption along with the use of live migration of VMs to minimize the number of active physical hosts. Thus, for energy efficient VM consolidation process first we proposed a median based threshold approach (MEDTH) for the selection of underutilized and over-utilized host present inside data center. This proposed approach has been validated and showed better SLA performance than other policies by consuming more energy. Since the objective of energy consumption has not been achieved in this step therefore, in second step we propose an Analytic Hierarchy Process a multi-criterion based approach for the selection of VMs for migration. This algorithm selects the VMs on the basis of their memory occupancy, CPU utilization and migration time to achieve the objective of energy minimization. Results obtained from the evaluation of proposed algorithm showed better energy savings on comparison to conventional techniques.

We have formulated the problem of dynamic VM Placement as a bin packing problem for the selection of most appropriate host. For which we have proposed three different heuristics based on best fit algorithms and they considered current and previous usage of the cloud resources. As, the main objective of this thesis is to minimize the energy consumption of data center but the excess of consolidation and migration during the process of VM consolidation may degrade the performance of system therefore, our solution focuses the overall performance of the system instead of only energy consumption to maintain a trust level between cloud's users and providers. Furthermore, we have also solved the problem of VM placement using nature inspired genetic algorithm and analyzed that how the use of genetic algorithm delivers the better results for VM placement than classical bin packing algorithms. The only disadvantage of these genetic algorithms is that they take more execution time for solving the problems because of their large search space. For, further improvements of energy savings by using the idea of genetic algorithms we have again formulated the problem of VM placement as multi objective optimization using Non-Dominated Sorting Genetic Algorithm (NSGA). Where, the overall objective is to minimize the energy and SLA performance along with the migration count in order to improve the performance of data center. Finally, to solve the problem of VM placement as multi-objective optimization we have designed a BPGA (Back Propagation-Genetic Algorithm) model which made use of both NSGA and back propagation neural network (BPNN). This BPGA model has been validated and it shows better results for energy consumption as well as other performance parameters of data centers. Thus, BPGA provides better energy performance trade-offs.

The proposed algorithms, heuristics and models for dynamic VM consolidation has been evaluated using an open source software framework called CloudSim using PlanetLab dataset for different workload conditions. Our solutions are robust, flexible and generic. Moreover, the experimental results shows that the overall VM consolidation process using the proposed BPGA VM Placement provides 19.8%, 9.7% and 4.5% of energy savings in comparison to GA, ACO and NSGA VM placement techniques

TABLE OF CONTENT

ABSTRACT……i-ii
1. INTRODUCTION 2
1.1 Motivation and contribution 2
2. OBJECTIVES 3
2.1 MEDIAN BASED THRESHOLD APPROACH FOR HOST SELECTION (MEDTH) 4
2.1.1 System model used for proposed work 4
2.1.2 Power model 5
2.1.3 Energy model 5
2.1.4 Performance Metrics 5
2.1.5 Experiments and Results 7
2.2 PROPOSED APPROACH FOR VM SELECTION USING AHP PROCESS 8
2.2.1 Experimental Results 9
2.3 ARBF- (AVAILABLE RESOURCE BEST FIT) SLA AND PERFORMANCE EFFICIENT HEURISTICS FOR VM PLACEMENT 10
2.3.1 Results and Discussion 10
2.4 META HEAURISTICS APPROACH FOR VM PLACEMENT (NSGA-VMP) 11
2.4.1 Problem formulation for NSGA-VMP 12
2.4.2 Experimental Results and discussions 13
2.5 BPGA-VMP: BACK PROPAGATION NEURAL NETWORK ALONG WITH GENETIC ALGORITHM VIRTUAL MACHINE PLACEMENT 14
2.5.1 Results and Discussion 15
3. CONCLUSION 16
4. REFERENCES 16
5. LIST OF PUBLICATIONS 19

1. INTRODUCTION

Proliferation of cloud computing is capable of supporting various computing services such as storage, servers, networks and applications for both e-sciences, e-business and many more over the network. This new paradigm of cloud computing is a large pool of readily usable and easily accessible virtualized resources that are provisioned as pay per usage model, where availability and guarantee of resources are offered by cloud service providers by the means of Service Level Agreement (SLA) [1]. Large numbers of data centers are required to respond the demands of customers, which results the consumption of huge amount of energy. From the previous research, it has been found that 55% of energy consumption of data center is due to servers, data equipment's and 30% is due to cooling equipment's therefore, these datacenters are not only expensive to maintain but they also have adverse effects on the environment due to huge amount of energy consumption and carbon emission. Reasons for high energy consumption are not only the power inefficient hardware and large number of computing resources but also the inefficient usage of these resources. Virtualization is the most appropriate topic in cloud computing, which provides better QoS and deals with auto scaling, server and VM consolidation, load balancing and many more. The objective of virtualization is to run multiple operating systems simultaneously on single physical machine by sharing the hardware resources. But the adoption of virtualization can also degrade the performance in terms of the percentage of CPU allocation to the VMs and on execution time [2]. Also, the improper allocation of VMs on unsuitable host affects the interference of the different applications on same physical machines and this leads to the performance degradation with decreased level of Quality of services (QoS) for the applications. Therefore, certain issues should be resolved during the adoption of virtualization process which can improve resource utilization, performance and energy consumption of data center.

1.1 Motivation and contribution

For the improvement of energy inefficient data centers, it is very important to understand how the power is distributed among the various components present inside data center. According to the report of EPA on data center's energy [3], the power consumption of a server is 40% of the

data center's power and 80% of the total IT load. Similarly, the survey from open compute project [4] reported that 91% of data center's energy consumption is only due to its computing resources. Thus, the source of high emission of CO_2 is also due to inefficient usage of computing resources. Data provided in [5] shows that most of the times, server's works more than 15-50% of their overall capacity which results over provisioning of the resources and leads to the higher Total Cost of Acquisition (TCA) [6]. According to National Resource Defense Council's report [9], [10], mostly the data centers are unused and the underutilization of servers is also inefficient for energy aware data centers. Solution of this problem is the consolidation of servers or VM consolidation by which fewer numbers of hosts will run the same application with lesser power consumption and it will further reduce the overall energy. Along with this, during the low server utilization, idle servers consume 70% of the power. Therefore, these idle servers should be turn off to reduce the energy consumption.

Thus, to solve the problem of energy efficient data centers, our contributions are described here by the adoption of VM consolidation. Consolidation of VMs is an effective technique for the reduction of energy and power consumption in the data center. Virtual machines are consolidated over physical nodes in such a way that a lesser number of physical nodes should be used and for efficient consolidation it would be useful to keep the idle nodes into low power state (sleep or hibernate) or turn them off. By doing this we can reduce the idle power consumption of data center. The problem of dynamic VM consolidation can be processed into different steps [11] as mentioned below and we have proposed novel methods for all the steps to achieve our objectives:

1. Detection of the overloaded and under loaded host present inside a data center which requires migration of some of its VMs.
2. Selection of VMs for migration, present inside this particular overloaded host machine.
3. Placement of the selected VMs from overloaded or under loaded host to some new host (VM Placement).

2. OBJECTIVES

To deals with the research issues that occur during dynamic virtual machine consolidation in IaaS cloud environment the problem of VM consolidation has been studied in the literature, but still there is a lack of optimal algorithms to ensure energy efficient VM consolidation. The

various problems are investigated to solve the problem of VM consolidation and our first objective is to find the over utilized and under-utilized host machines with in the data centers to start the process of consolidation.

2.1 MEDIAN BASED THRESHOLD APPROACH FOR HOST SELECTION (MEDTH)

MEDTH method start with the calculation of the CPU utilization of all physical machines present inside datacenter in the first step. After this, the upper and lower threshold values are to be calculated such that these threshold values will be further used for finding the hot spot i.e. over utilized physical machine and excess capacity server i.e. under-utilized physical machine. For the CPU utilization of all physical machines, we have used random generator. C_i represents the CPU utilization of P_i numbers of host machines, where i ∈ R+ and P_i number of host machines can be arranged as: {P_i = P_1, P_2, P_3 ...P_i}. These physical machines can be even and odd in numbers, therefore median of even numbers of machines i.e. P_{2i} gives two new sets: first set is (P_1 to P_i represented as X_i), and second set is (P_i +1 to P_{2i} represented as X_j). Similarly, two different sets can also be arranged for odd numbers of physical machines i.e. P_{2i} +1. First set is (P_1 to P_i represented as Y_i) and second set is (P_i +2 to P_{2i} +1 represented as Y_j). Similarly, upper and lower threshold limits Th_u and Th_l for both even and odd number of physical machines present inside data centers can be detected using median method. With their help, the over utilized and under-utilized physical machines can be found as following:

$$\begin{cases} if\ CPUP_i > Th_u\ (P_i = Over_{host} \\ if\ CPUP_i < Th_l\ (P_i = Under_{host} \end{cases} \quad (1)$$

Here in above equation 1, $Over_{host}$ represents over utilized hosts and $Under_{host}$ represents under-utilized physical machines. If CPU utilization of physical machine is greater than upper threshold value then machine will be consider as over utilized and if the CPU utilization of machine is lesser than the value of lower threshold then it will be considered as under-utilized machine.

2.1.1 System model used for proposed work

The system model of our work contains N numbers of heterogeneous host machines. Each host machine is characterized with CPU performance in terms of MIPS, network bandwidth and amount of RAM occupied by that particular host machine. These host machines are composed of several heterogeneous virtual machines to run user applications. These virtual machines are also characterized by MIPS, Network bandwidth and RAM, so that multiple user requests for VM

provisioning with their particular characteristics. It contains two important parts: 1) Global manager 2) Local manager. Global manager acts a resource manager inside data centers and allocates VMs to available host machines on the basis of predefined characteristics whereas, local manager acts as a decision maker for deciding which and when a VM should be migrated to another host. SLA has been established between the cloud users and the service providers for the delivery of predefined QOS to the user and in the case of SLA violation, a provider has to pay penalty to cloud user

2.1.2 Power model

CPU performance, memory, disk storage, power supplies and cooling systems are the important factors for determining power consumption of servers in the data centers. As the CPU is the biggest source of power consumption inside the data centers and there is a linear relationship between CPU utilization and power consumption. We have used the data for the power consumption of hosts, which have been provided by SPEC power benchmark [10]. We have used six different servers in our work and these servers are: HPProLiant ML110 G4, HPProLiant ML110 G5, IBM Server x3250XeonX3480, IBM Server x3550XeonX3470, Acer AT150 F1, Acer AR320 F1.

2.1.3 Energy model

Energy consumption can be calculated with the help of this model. As minimization of energy consumption is the main aim of current research, therefore it is very important to explain the model of energy consumption. This energy consumption can be calculated as the summation of total power consumption of host during every period of time, explained in equation 2.

$$E(t) = \int P(t)dt \qquad (2)$$

2.1.4 Performance Metrics

For measuring the performance of cloud environment, it is very important to analyze the performance metrics during the proposal of models. In our work, we have considered the QOS as a performance metrics which are formalized in SLA- Service Level Agreement that has been signed between the cloud users and providers. These QOS can be determined in terms of characteristics and can be varied for different applications; therefore, it is very important to define such metric to calculate SLA delivered. In our work, we use following performance

metrics: SLATAH (SLA violation time per active host), PDM (performance degradation due to migration) and SLAV (SLA violation). Multiplication of these two metrics SLATAH and PDM introduces SLA violation i.e. SLAV defined in equation 5 and our focus is to minimize this SLAV metric for meeting QOS**.**

1. **SLATAH** (SLA violation per active host): percentage of time when CPU utilization of host machine reaches to 100%.

$$SLATAH = \sum_{i=1}^{N} \frac{T_{si}}{T_{ai}} \quad (3)$$

In equation 3 N represents the number of hosts, whereas T_{si} is the time when CPU utilization reaches to 100% and T_{ai} is the total time for which host remains active.

2. **PDM** (Performance degradation due to migration): calculates total percentage of performance degradation during VM migration. Author in [8] mentioned that this performance degradation of applications depends upon the host CPU utilization

$$PDM = \frac{1}{M} \sum_{j=1}^{M} \frac{C_{dj}}{C_{rj}} \quad (4)$$

In equation 4 M represents the number of VMs, C_{dj} estimates the performance degradation of VMj during VM migration. C_{dj} can be estimated as 10% of CPU utilization during all VMj migration. C_{rj} is total CPU capacity requested by VMj. These two metrics are equally important for SLA violation therefore, another metric SLAV is the combination of both SLATAH and PDM shown in equation 4

3. **SLAV** (Service Level Agreement Violation): The above two metrics are equally important for SLA violation therefore, SLAV is the combination of both SLATAH and PDM shown in equation 5

$$SLAV = SLATAH * PDM \quad (5)$$

4. **ESM** (Energy SLA Migration): the objective is to minimize both energy and SLA therefore; we have combined energy (E) and SLA (S). Moreover, excess of migration may degrade the performance of system therefore we have also considered migration count (M). Thus, the combined metric ESM has been evaluated in our experiments for analyzing the comparison among existing algorithms used during VM consolidation process shown in equation 6.

$$ESM = Energy * SLAV * Number\ of\ migration \quad (6)$$

5. **Cost of data center:** the service provider's needs to pay the cost of energy that is consumed by physical machines. Therefore, cost function is also important for the consideration during

live migration and this cost function can be calculated as shown in equation 7 where CTP is the cost of power and tp is time period (energy per unit time) and CTV is the cost of SLA violation per unit time. SLA violation between service provider and consumer occurs only if the demand of resources exceeds the available capacity of resources. In equation 7, *to* is the initial time frame and T is the total time. HAtj indicating whether host j is active on not at the time t and similarly, SVtj indicating whether host j is experiencing SLA violation or not in time frame t.

$$CT = \sum_{t=t0}^{T}(CTP \sum_{j=0}^{M} HA_{tj} + CTV \sum_{j=1}^{M} SV_{tj}) \quad (7)$$

2.1.5 Experiments and Results

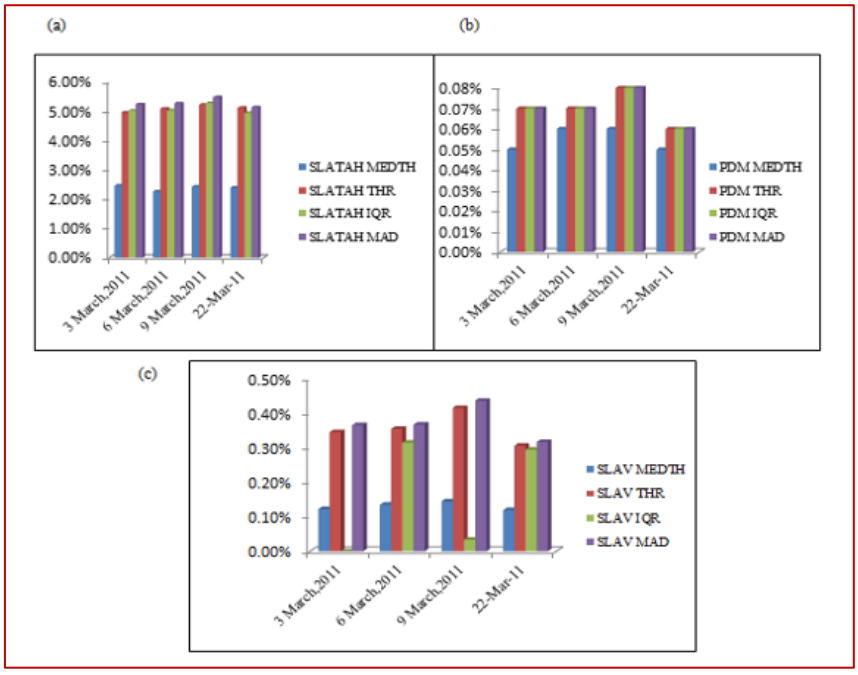


Figure 1: (a) Percentage of SLATAH, (b) Percentage of SLAV and (c) Percentage of PDM for real workload

For the performance evaluation of proposed approach, we have considered the real workload environment. For real workload, we have taken the data from CoMon project, a monitoring infrastructure of PlanetLab [11]. This project includes the data for CPU utilization of thousands of VMs which is obtained from servers located more than 500 places around the world and data is collected after every five minutes. We have selected data of four days from workload traces of the project during March2011.There are different numbers of VMs for each day, from which

each VM is randomly assigned a workload trace for every corresponding day. The simulation has been conducted in CloudSim toolkit [12] and figure 1 shows the results.

2.2 PROPOSED APPROACH FOR VM SELECTION USING AHP PROCESS

For the selection of VMs we have proposed a novel technique using multi-criteria decision-making process i.e. AHP- Analytic Hierarch Process. Multi criteria decision making is one of the most well-known branches of decision making which is associated with multiple attributes. Our proposed method considers three resources CPU utilization, memory occupancy and migration time simultaneously for VM selection. It selects the VM by taking the advantage of different criteria's using three resources. As there are three different criteria's for selection, therefore we need to choose most dominating among all. Analytic hierarchy process uses relative values rather than actual. The process of selecting best alternative starts by creating an N * M matrix where N is the number of alternatives of VM and M represents number of criteria and shown below.

$$VM_{AHP} = \begin{bmatrix} C_{vm1} & MO_{vm1} & MT_{vm1} \\ C_{vm2} & MO_{vm2} & MT_{vm2} \\ C_{vm_n} & MO_{vm_n} & MT_{vm_n} \end{bmatrix} \quad (8)$$

Here vm1, vm2,..... vm_n is the available virtual machines present inside the selected host machine and C, MO, MT are three different criteria representing CPU utilization, memory occupied by VM and migration time. In order to select the best alternative of VM we go through different steps as follows:

Step 1: First, we normalize the matrix VM_{AHP} dividing each entry by the total values of each column as shown in equation 9.

$$Score(q_{ij}) = \begin{bmatrix} \frac{C_{vm1}}{C_{total}} & \frac{MO_{vm1}}{MO_{total}} & \frac{MT_{vm1}}{MT_{toal}} \\ \frac{C_{vm2}}{C_{total}} & \frac{MO_{vm2}}{MO_{total}} & \frac{MT_{vm2}}{MT_{total}} \\ \frac{C_{vm_n}}{C_{total}} & \frac{MO_{vm_n}}{MO_{total}} & \frac{MT_{vm_n}}{MT_{total}} \end{bmatrix} \quad (9)$$

Step 2: In this step, the better alternative can be selected from equation 10 by multiplying the score obtained from equation 9 with the predefined weights, where q_{ij} is the score of matrix and w_{ij} is the weight of each criterion. In this study, we consider all the resources equally important therefore we have assigned equal weights to all the criteria's.

$$VM_{AHPnew} = max_i \sum_{j=1}^{M} q_{ij} * W_{ij} \ (for\ i = 1,2,3 \dots \dots \dots N) \quad (10)$$

Step 3: In the last step, VMs are ranked according to VM_{AHPnew} values and VM with the highest score is selected for migration.

2.2.1 Experimental Results

For the evaluation of proposed policy in real environment, ten experiments are executed for the 10 days of workloads and then their average results for energy consumption, SLA violation, ESV (combined metric of Energy and SLAV) and total execution time to complete VM consolidation process. We have compared the AHP VM selection method with rest of the four existing VM selection policies such as MU (Maximum utilization), MC (Minimum Correlation), MMT (Minimum Migration Time) and RS (Random Selection) VM selection methods.

Median Absolute Deviation (MAD) as a host overload detection algorithm and results are calculated using MAD with five different VM selection policies.

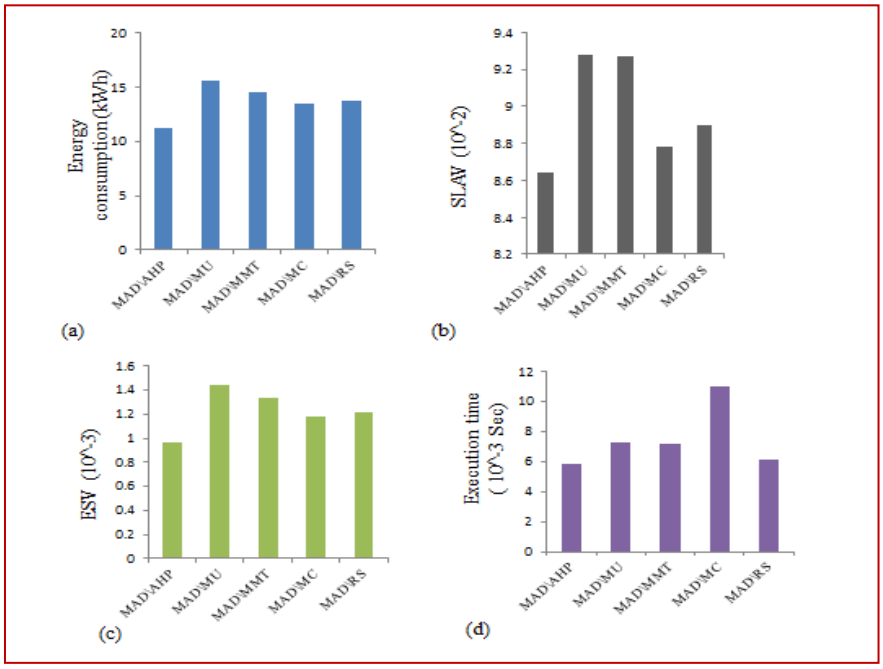


Figure 2: Graphs of VM selection policies for different parameters (a) Energy consumption using MAD with different VM selection policies. (b) SLA violation for various selection policies. (c) ESV metric and (d) Execution time for VM consolidation process by using various VM selection policies with MAD

Here, we have considered MAD - Median absolute deviation as host overload detection policy and calculate its impact along with existing VM selection policies as well as our proposed method. Figure 2 shows the results for MAD\AHP, MAD\MU, MAD\MMT, MAD\MC and MAD\RS.

2.3 ARBF- (AVAILABLE RESOURCE BEST FIT) SLA AND PERFORMANCE EFFICIENT HEURISTICS FOR VM PLACEMENT

To solve the problem of VM placement using the basics of bin packing here, we have proposed three different heuristics by considering three types of computing resources such as physical memory (RAM), MIPS and the power consumption of each host machine during VM placement These three heuristics are also the modifications of BFD- Best Fit Decreasing and named as: (ARBFH1) Available Resources Best Fit Heuristic 1, ARBFH2 (heuristic 2) and ARBFH3 (heuristic 3) for performance efficient VM Placement. VM List stores the VMs that needed to be migrated. VMs can be mapped according to available resources and utilization of past resources using ARBF strategy.

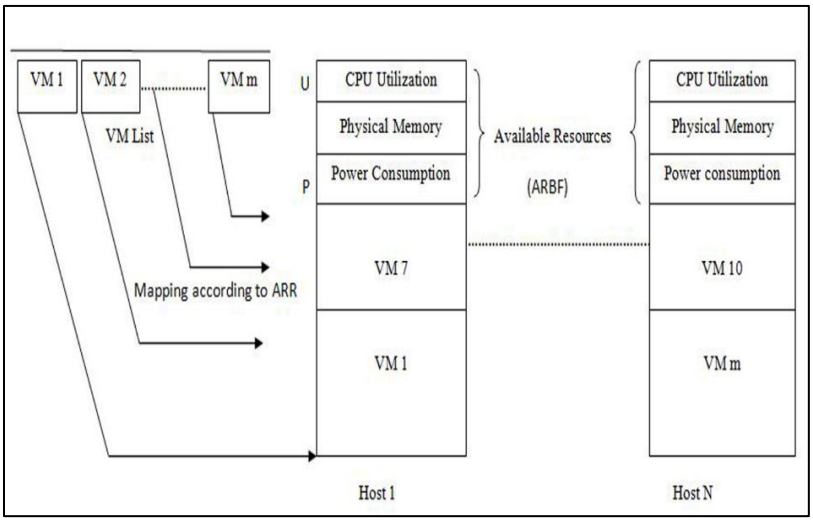


Figure 3: VM placement model using Available Resources Best Fit (ARBF)

2.3.1 Results and Discussion

The process of VM consolidation inside data centers includes different phases. In this section, the combination of best policies reported by Anton et.al [9] are considered as a reference scenario for the comparison. LR (Linear regression), MMT (Minimum migration time) and PABFD is compared to the scenario described in [13] and our proposed heuristics to find out the combination of best policy. Five different performance factors affect from these policies: 1) SLA violation, 2) number of migrations and Energy consumption and ESM for the overall consolidation process. But in order to compare the net effect of each policy ESM metric has been evaluated and used to find out the best one.

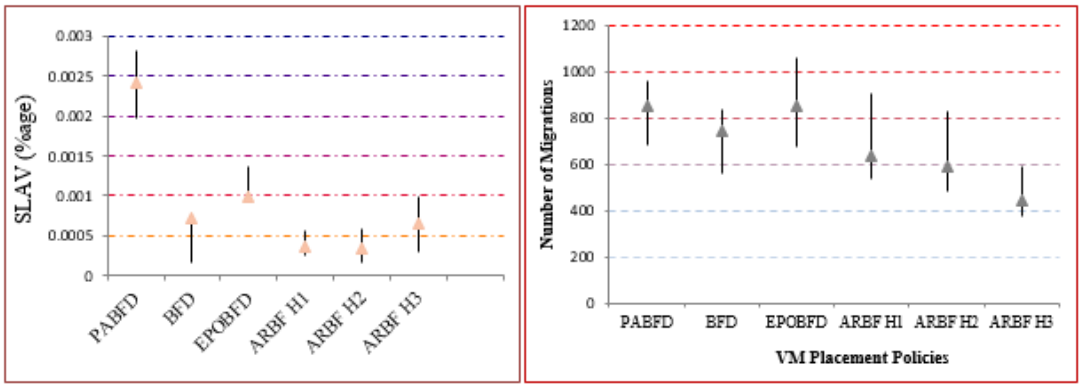


Figure 3: SLA Violation and number of migration using various heuristics of VM Placement

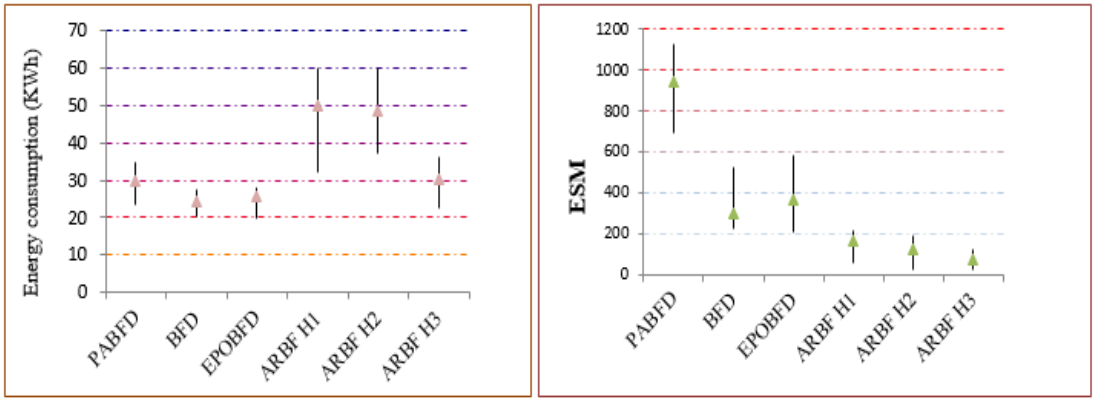


Figure 4: Results of Energy Consumption and ESM for different heuristics used for VM placement

As depicted in Figures 3, the adoption of ARBF heuristics H1, H2 and H3 provide better SLA violation and Number of migration during VM placement. However, these heuristics increase the energy consumption of data centers but simultaneously provides improvement in the results of ESM metric with minimizing the level of SLA violations and number of migrations during simulation shown in figure 4. As, here our objective was to deliver the better QoS and performance efficient SLA to the cloud users's without the consideration of energy consumption therefore, the increase in energy consumption can be inconsiderable.

2.4 META HEAURISTICS APPROACH FOR VM PLACEMENT (NSGA-VMP)

This section highlights the proposed solution of VM placement using NSGA – Non-dominated Sorting Genetic Algorithm. As, in single objective optimization problem, only one attempt need to be made for finding best solution but in case of multi-objective optimization problem there is a set of solutions available in the search space. However, some of them are superior when all the

objectives are considered but inferior during the consideration of one or more objectives and these solution sets are known as pare to optimal solutions or non-dominated solutions. Therefore, NSGA has been proposed to obtain the optimal solution in such cases

2.4.1 Problem formulation for NSGA-VMP

To solve the problem of virtual machine placement as a multi objective optimization problem, the focus of current work is to find an optimal solution for the placement for which system considerations of data centers are defined as follows:

Vm: a set of virtual machines.
Pm: a set of physical machines.
Vm_iCPU: CPU requirements of virtual machines (from i=1 to n).
Vm_iMem: memory requirements of virtual machines (from i= 1 to n).
Vm_iBw: bandwidth requirements of virtual machines (from i= 1 to n).
Pm_jCPU: CPU capacity of physical machines (from j=1 to m).
Pm_jMem: memory of physical machines (from j=1 to m).
Pm_jBw: bandwidth of physical machines (from j=1 to m).
$VmPm_j$: entire set of virtual machines that are allocated to physical machines 1 to m.

Therefore, Virtual machine 1 can be described as *Vm_1 (CPU, Mem, Bw*) and similarly physical machine can be described as *PM_1 (CPU, Mem, Bw)*. The aim of this investigation is to find an optimal mapping of Virtual machines over physical machines that satisfy the requirements of virtual machines while achieving the following three objectives:

Objective1: Minimizing the energy consumption (E)
Objective 2: Minimizing the level of SLA violation (S)
Objective 3: Minimizing the number of VM migrations (M_c)

Overall objective function

Every individual that has been generated in new population have to fulfill all three objectives. Therefore, the value of final objective function be calculated using Equation 11. Individual with minimum value of objective will provide optimal solution for the placement

$$objective = Minimize(\ E * S * M_C) \tag{11}$$

The proposed algorithm works as follows: population is generated randomly with set of random VM-PM mapping. If population is classified then, genetic operator's crossover and mutation will be applied on the population and population will be further evaluated to check whether it fulfill the objective Function or not. On the other side, if population is not classified then NSGA will be

performed before the selection operator work. In this, first non-dominated individuals will be identified from the population and ranked according to their non-domination level. Individuals of first non-dominated front will be assigned a large dummy fitness value according to first front of individuals will be created. Similarly, fronts will be generated and process of evolution goes on by increasing the front and generation counter. Same steps will be followed for evolutionary process until the algorithm meets the stopping criteria.

2.4.2 Experimental Results and discussions

The results analysis of proposed NSGA- VM placement has been implemented in JAVA Net beans IDE using CloudSim toolkit, which is an existing cloud simulation framework. We have taken the crossover and mutation rate as 0.5 the validation of proposed algorithm along with single point crossover. Population size has been taken as 500. For the simulation inside cloud environment, the numbers of physical machines and virtual machines used are different according to the test cases. These test cases vary in terms of server configuration as well as number of virtual machines and number of physical machines. Different instances of VMs are taken from Amazon EC2 [21]. Here, we have used five different of VMs we have used.

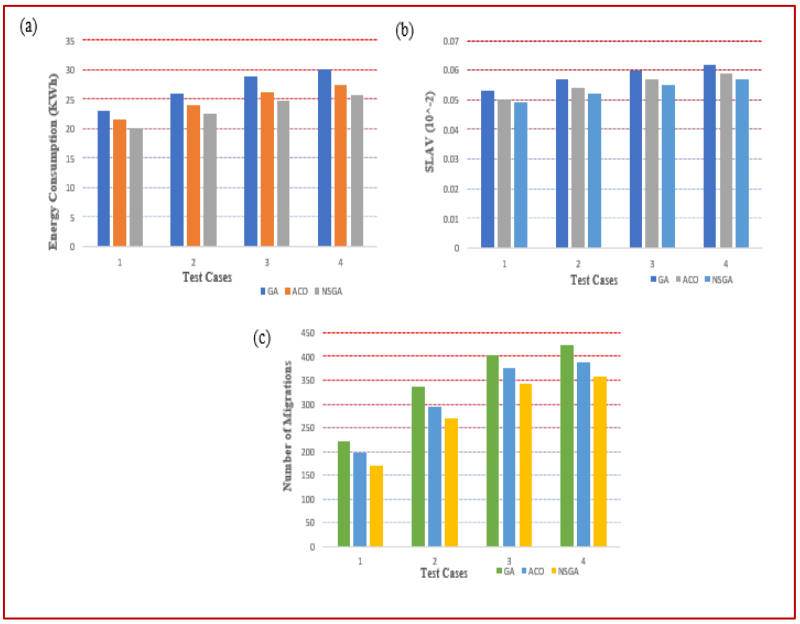


Figure 10: (a) Energy consumption for NSGA, ACO and GA (b) SLAV using NSGA, ACO and GA (c) Number of Migrations taken during NSGA, ACO and GA

Figure 10 provides the results for energy consumption, SLA violation and migration count. We have compared the results with basic GA (Genetic Algorithm) and ACO (Ant Colony Optimization) approach of VM Placement. The reason for taking lesser number of migration during VM placement using NSGA is that, the NSGA already rank the population in different fronts during each iteration and their ranking is based on the sorting of their non-domination level. Thus, the creation of fronts in this approach will reduce the chance of migration for the placement of VM over PM. Since, the migration count is less therefore; performance degradation during migration also will be less. Moreover, the hosts will remain active for longer time due to creation of fronts again and again, which leads to lesser SLA time per active host. As, the SLA violation is the combination of SLATAH and PDM therefore; this method will provide minimum SLA violation during placement. Due to minimum resources used in this method, minimum will be the power consumption per unit of time and minimum will be the energy consumption.

2.5 BPGA-VMP: BACK PROPAGATION NEURAL NETWORK ALONG WITH GENETIC ALGORITHM VIRTUAL MACHINE PLACEMENT

First three steps of VM consolidation process has been already done in our previous work where we have proposed an energy aware algorithm for the selection of VMs for migration and selection of underutilized, over utilized host. Now, here we have proposed BPGA model which facilitates the energy aware VM placement algorithm to minimize the energy consumption of datacenters without degrading the performance of datacenter. Therefore, by increasing the utilization of server's resources and reducing the number of active servers minimize the energy consumption and thus, directly contributing to the green computing. The current work presents an energy aware virtual machine placement (VMP) technique for cloud datacenters which is based on genetic algorithm (NSGA) and biological neural network (BPNN) and named as BPGA-VMP model. This BPGA-VMP model is working in two passes. In first pass the process of Non-dominated sorting Genetic Algorithm (NSGA) will be used for the mapping of virtual machines. NSGA contains two objective functions and the physical machine which fulfill all the requirements through NSGA would be selected for the mapping of VMs and rest of the VMS are forwarded to the second pass of BPGA model which follows the process of training the neural network based on the feature of power consumption to decide which physical machine should be used for particular VM.

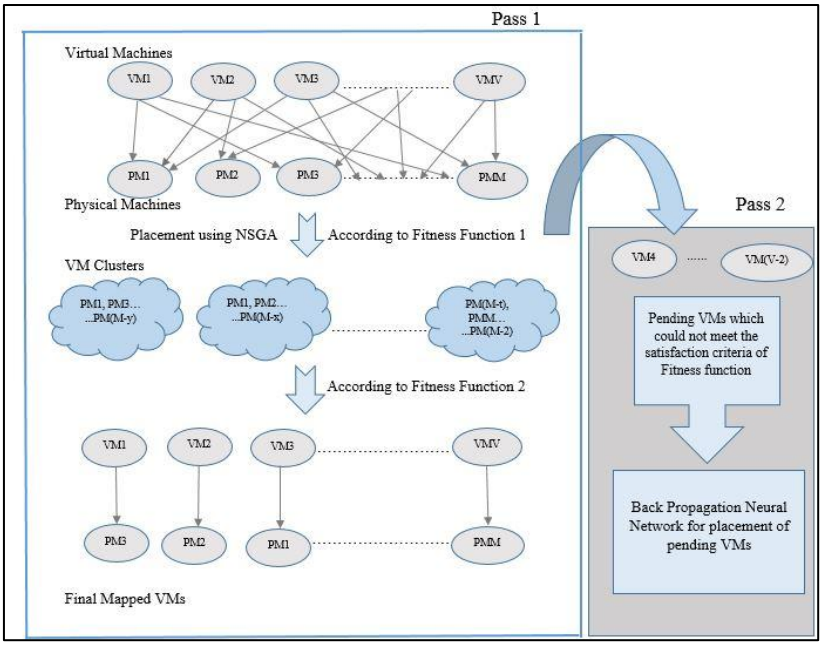


Figure 11: BPGA-VMP model for VM placement

2.5.1 Results and Discussion

800 hosts have been used for the simulations, whereas number of runs for the simulation are 10 with different number of VMs on every run (with different workload condition).

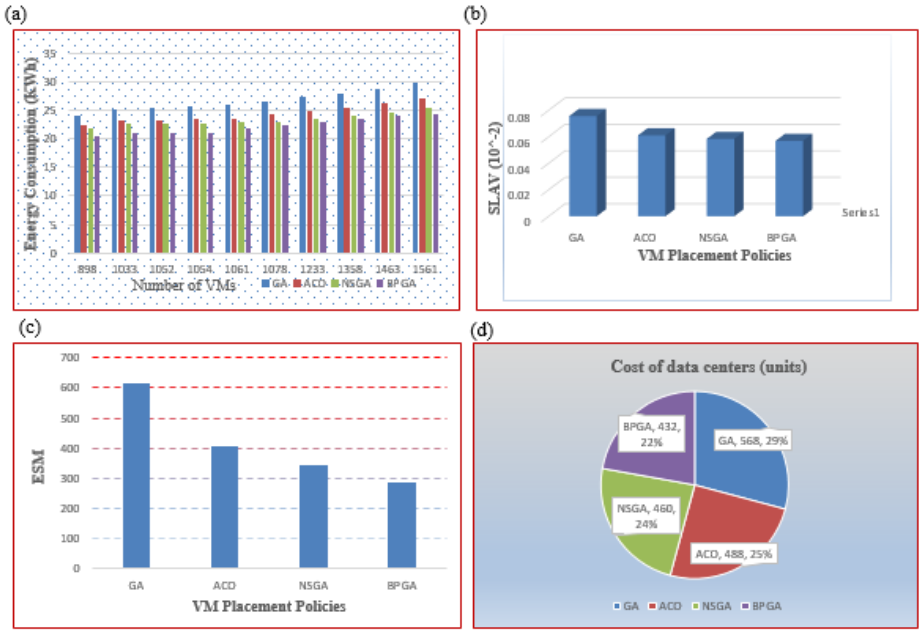


Figure 12: (A) Energy Consumption V/s VM Placement Techniques, (B) SLA Violations V/s VM Placement Techniques, (C) ESV V/s VM Placement Techniques and (D) Cost of data center using various VM Placement Techniques

Total energy consumption, number of host used after placement, migration count, SLA Violations occurs during VM placement and Cost of data center for overall process have been calculated through proposed BPGA-VMP model. The obtained results have been compared with GA [15], [16], [17], ACO [18], [19], [21] and NSGA based techniques for VM placement. An average of 19.8 %, 9.58% and 4.5% of energy has been saved by our proposed BPGA-VMP Placement technique in comparison to GA, ACO and NSGA respectively

3. CONCLUSION

The working environment of IT industry has revolutionized the trend of utility based computing i.e. cloud computing. Due to this increasing trend, the demand for cloud data centers are increasing day by day which consumes a large number of world's computing resources and results a huge amount of energy consumption. Therefore, the efficient management of data center's resources and CO_2 emission because of high energy consumption are two important issues to handle for environmental sustainability. The process of dynamic consolidation has been considered as one of the most effective methods for minimizing the energy consumption of data centers. Therefore, this thesis has presented different novel algorithms or methods for dynamic VM consolidation while dealing with QoS. Proposed approaches minimize the energy consumption or we can say that provide a tradeoff values for energy consumption as well as SLA violation while dealing with other QoS metrics. Also, our proposed model minimizes the cost of the data center and all others mentioned objectives.

Future work is planned to evaluate the proposed models or algorithms in Cloud Stack a real infrastructure for the cloud environment. Moreover, the direction for future research will be implementing energy aware resource allocation algorithms using different configurations of data centers such as network topology, cooling structure and many others.

4. REFERENCES

[1] M.H. Ferdaus, & Murshed. Energy-aware Virtual Machine Consolidation in IaaS Cloud Computing, *Cloud computing*, Springer International Publishing, pp. 179-208, 2014.

[2] G. Kousiouris, T. Cucinotta, & T. Varvarigou, "The effects of scheduling, workload type and consolidation scenarios on virtual machine performance and their prediction through optimized artificial neural networks", *Journal of Systems and Software*, vol. 84, no. 8, pp. 1270-1291, 2011.

[3] Report to congress on server and data center energy efficiency, environmental protection agency. Available at www.energystar.gov/ia/partners/prod_development/downloads/EPA__Datacenter_Report_Congress_Final1. pdf. Retrieved on 17/10/2016.

[4] Open Compute Project, Energy efficiency, Available: http://opencompute.org/about/ energy-efficiency/. Retrieved on 21/11/2016.

[5] L. A. Barroso & U. Holzle, "The case for energy-proportional computing," *Computer*, vol. 40, no. 12, pp. 33–37, 2007.

[6] W. X. Fan, D. Weber & L. A. Barroso, "Power provisioning for a warehouse-sized computer," *Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA)*, pp. 13–23, 2007.

[7] Natural resources defense council. Available at http://www.nrdc.org/energy. Retrieved on 18/1/2017.

[8] Scaling up energy efficiency across the data center industry: evaluating key drivers and barriers, NRDC. Available at: http://www.nrdc.org/energy/files/data-center-efficiency-assessment-IP.pdf. Retrieved on 20/1/2017.

[9] A. Beloglazov & R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers", *Concurrency and Computation: Practice and Experience*, vol. 24, no. 13, pp. 1397-1420, 2012.

[10] SPEC power benchmarks, Standard Performance Evaluation Corporation. Available at http://www.spec.org/benchmarks html#power. Retrieved on 15/5/2016.

[11] K. S. Park & V. S. Pai, "CoMon: a mostly-scalable monitoring system for Planet-Lab," *ACM SIGOPS Operating Systems Review*, vol. 40, no. 1, pp. 65–74, 2006.

[12] N. Rodrigo, Calheiros, R. Ranjan, A. Beloglazov, A. F. Cesar, D. Rose & R. Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", *In Software Practice and Experience*, vol. 41, pp. 23-50, 2011.

[13] Q. H. Nguyen, T. Nam & T. Nguyen, "EPOBF: Energy Efficient Allocation of Virtual Machines", *High performance computing*, vol. 51, no. 4B, pp. 173-182, 2013.

[14] W. Voorsluys, J. Broberg, S. Venugopal & R. Buyya, "Cost of virtual machine live migration in clouds: A performance evaluation", *Proceedings of Ist International Conference on Cloud Computing*, pp. 254-265, 2009.

[15] B. Madhusudhan & K. C. Sekaran, "A genetic algorithm approach for virtual machine placement in cloud. *Proceedings of International Conference on Emerging Research in Computing, Information, Communication and Application (ERCICA),* 2013.

[16] F. F. Moghaddam, R. F. Moghaddam & M. Cheriet, "Carbon –aware distributed cloud: multi-level grouping genetic algorithm", *Cluster computing*, vol. 18, no. 1, pp. 477-491, 2015.

[17] Y. S. Dong, G. C. Xu & X. D. Fu, "A distributed parallel genetic algorithm of placement strategy for virtual machines deployment on cloud platform", *The scientific world journal,* Article ID 250139, 2014.

[18] M. A. Tawfeek, A. B. Elsisi, A. E. Keshk & F. A. Torkey, "Virtual machine placement based on ant colony optimization for minimizing resource wastage", *Proceedings of Advanced Machine Learning Technologies and Applications*, pp. 153–164, 2014.

[19] A. Layeb & Z. Benayad, "A novel firefly algorithm based ant colony optimization for solving combinatorial optimization problems", *International Journal of Computer Science and Applications*, vol. 11, no. 2, 2014.

[20] Amazon elastic computing cloud (EC2), Available at http://aws.amazon.com/ec2/instance-types. Retrieved on 11/2/2016.

[21] N. J. Kansal & I. Chana, "Energy-aware Virtual Machine Migration for Cloud Computing-A firefly Optimization", *Journal of Grid Computing*, vol. 14, no. 2, pp. 327-345. 2016.

[22] Y. W. Foo, C. S. Goh and Y. Li, "Machine Learning with Sensitivity Analysis to Determine Key Factors Contributing to Energy Consumption in Cloud Data Centers", *Proceedings of International Conference on Cloud Computing Research and Innovation*, 2017.

5. LIST OF PUBLICATIONS

Published Journal Papers

- O. Sharma and H. Saini. "Energy Efficient Virtual Machine Consolidation for Cloud Data Centers Using Analytic Hierarchy Process". *In International Journal of Advanced Intelligence and Paradigms*. (In Press)
  [Major indexing: SCOPUS (Elsevier), DBLP, ACM Digital Library, Academic OneFile (Gale), Educational Research Abstracts]. *Online link:* http://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijaip

- O. Sharma and H. Saini. "Performance Evaluation of VM Placement Using Classical Bin Packing and Genetic Algorithm for Cloud Environment". *In International Journal of Business Data and Communication Network (IJBDCN)*, vol. 13, no.1, pp.45-57, 2016.
  [Major indexing: SCOPUS, ESCI, DBLP, ACM digital Library, web of sciences, Google scholar].

- O. Sharma and H. Saini. "SLA and Performance Efficient Heuristics for Virtual Machine Placement inside Cloud Data Centers". *In International Journal of Grid and High Performance computing (IJGHPC)*, vol. 9, no.3, pp. 17-33, 2017. *[Major indexing: SCOPUS, ESCI, DBLP, ACM digital Library, web of sciences, Google scholar]*.

- O. Sharma and H. Saini. "State of Art for Energy Efficient Resource Allocation for Green Cloud Data centers". *In International Journal of Control Theory and Application (IJCTA)* vol.9, no.11, pp. 5271-5280, 2016. *[Major indexing: SCOPUS, EBSCOhost]*.

Published Conference Proceedings

- O. Sharma and H. Saini. "Experimental Analysis for Energy Management Techniques for Mobile Devices Using Cloud Computing". *In the Proceeding of International Conference on Green Computing and Internet of Things (ICGCIoT)*, pp. 737-742, 2015. *[Major indexing: SCOPUS, Conference Proceedings Citation Index]*.

- O. Sharma and H. Saini. "VM Consolidation for Cloud Data Centers using Median Based Threshold Approach". *In the proceedings of 12th International Multi-Conference on Information Processing (IMCIP-2016)*, vol. 89, pp.27-33, 2016. *[Major indexing: SCOPUS, Conference Proceedings Citation Index]*.

Communicated Journal Papers

- O. Sharma and H. Saini. "Energy & SLA Efficient Virtual machine placement in Cloud Environment using NSGA (Non-dominated Sorting Genetic Algorithm)". *In International Journal of Information Security and Privacy (IJISP)*. *[Major indexing: SCOPUS, ESCI, DBLP, ACM digital Library, web of sciences, Google scholar]*.

- O. Sharma and H. Saini. "BPGA: A Novel Approach for Energy Efficient Virtual Machine Placement in Cloud Data Centers". *In Journal of Computing (JOC). [Major indexing: SCI, SCIE, SCOPUS, ACM digital Library, web of sciences, Google scholar]*.

Communicated Book Chapter

- O. Sharma and H. Saini. "Performance Evaluation for Energy Aware VM Placement Techniques for Cloud Data Centers". In Advances in Data Communications and Networking for Digital Business Transformation. *IGI Global*. *[Major indexing: SCOPUS, Google Scholar, DBLP]*

Name and Signature of Candidate with date:

Name and Signature of Supervisor with date: